Generating more realistic complex networks from power-law distribution of fitness

G. A. Mendes and L. R. da Silva

Departamento de Física and National Institute of Science and Technology for Complex Systems,
Universidade Federal do Rio Grande do Norte
59072-970 Natal-RN, Brazil

In this work we analyze the implications of using a power law distribution of vertice's quality in the growth dynamics of a network studied by Bianconi and Barabási. Using this suggested distribution we show the degree distribution interpolates the Barabási *et al.* model and Bianconi *et al.* model. This modified model (with power law distribution) can help us understand the evolution of complex systems. Additionally, we determine the exponent gamma related to the degree distribution, the time evolution of the average number of links, $< k_i > \propto (t/i)^\beta$ (i coincides with the input-time of the ith node), the average path length and the clustering coefficient.

Keywords: Power-law, Scale-free networks, Complex networks, Fitness model

1. INTRODUCTION

The complex system's range is broad, covering not only the daily phenomenon in our lives, but sophisticated problems with amazing results (phase transition, S.O.C, small-world). One subject of this range is the complex networks that have received a lot of attention from researchers. It is composed by a set of N nodes (or vertices) and a set of links (or edges) that connect two vertices in the system. The large-scale topology systems are determined by in the dynamic forces that act at the level of an individual node. The preferential attachment might have important consequences in dynamic process taking place in the topology defined by the network because it is responsible for the system's correlations. Nowadays the study of complex networks is focused on random graphs and scale-free nets. For a long time graphs have been modeled as completely random. A well known example is the paper published by the Hungarian mathematicians Erdös and Rényi [1] where a network generated by placing links among the nodes at random is studied. However such a model cannot describe the topological properties of real complex networks due to some physic quantities them associated by usual exponential laws instead of power-law asymptotic behavior. Furthermore, in general real networks are more complex than Erdös and Rényi model and exhibit birth and death of nodes, aging [2], the links depending on some parameter (connectivity [3], fitness [4, 5], geographic distance [6, 7], etc).

In 1999 Barabási and Albert (BA) [3] introduced the model of random growth graphs displaying power law distributed connectivity. To obtain this fantastic result, they used at least two coexisting mechanisms, they are (i) growth (the system expands by adding new nodes which attach to the nodes already present in the network) and (ii) preferential attachment (a new node links to nodes which already exist. This process takes place with a higher probability when the sites present a higher number of neighbours "Barabási and Albert model"). Despite its success in exhibiting the degree distribution following a power law (Fig. 1), this simple and basic model predicts that all nodes increase their connectivity in time as $k_i(t) = (t/t_i)^\beta$, where $\beta = 1/2$ and t_i is the time at which node i has been added into the network. As the exponent has the same value for all nodes, the oldest vertices will have the highest number of links. Note that new nodes attach preferentially to the oldest ones, i.e. the most connected ones.
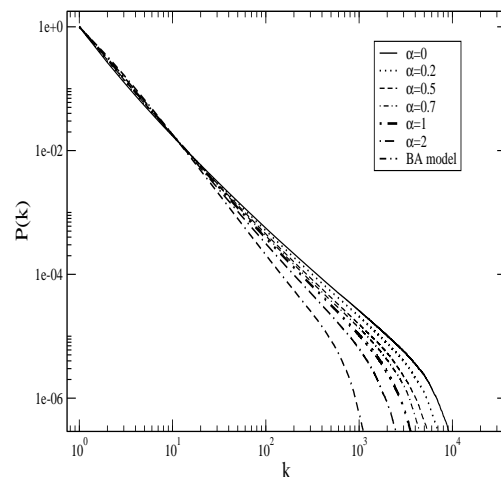

FIG. 1: It shows the cumulative degree distributions $P(k)$ in log-log plot for typical values α and BA model ($\gamma_{theoretic} = 3$). The continuous lines are the numerical results (network with 100000 nodes and 3000 runs).

Nowadays scientists have observed power law in a lot of systems as scale-free nets, systems with fractal geometry, percolation, diffusion-reorganized aggregates and others. All examples cited above indicate that power-law is very common and popular in the academic problems due to their uncountable applications in different fields of knowledge. Furthermore, recently there has been increasing evidence that real nets have some non-random characteristics, i.e., they obey various scaling laws, are fat-tailed and display short small length-scale clustering that is typical of scale-free structures [8–15]. For example we can cite the papers published in the last decade that have pointed out that many natural and social networks are scale-free. The degree distribution $P(k)$ gives the probability that randomly chosen node which has exactly k nearest neighbors behaves as $k^{-\gamma}$ for sufficient large degrees k. In general, the exponent γ lays between 2 and 4 [16, 17]. Other examples are: the World Wide Web where HTML documents are the nodes and the connections to other documents in the WWW are the edges [17]; the web of hu-

man sexual contacts where individuals are the nodes and sexual relations among them are the links [9]; and the citation networks where published articles are the nodes and a link represents a reference to a previously published article [10]. All these examples establish a methodology that defines links and vertices in their specific subject. However they appear to display considerable topological similarity, having degree distributions which behave as power-law.

Another feature easily seen from the examples above is the number of nodes increases during the evolution of the net. Better observing the WWW, we note that some new HTML documents acquire links at a higher rate than the oldest nodes (e.g. the site http://www.youtube.com has a high degree distribution despite being new). This phenomenon allows new documents to become hubs. Our intuition indicates that preferential attachment for systems similar to WWW must depend on connectivity and other parameters. A good candidate to be the other parameter is something associated to content and marketing of this document. Bianconi and Barabsi [4] associate the "good candidate" with some intrinsic quality of nodes, such as the content of a web page in the WWW, the skill in the sexual network or the knowledge in the citation network. They call this parameter the node's fitness. In this direction, the system becomes heterogeneous and its individual elements influence the net depending on the quality factor.

The plan of this paper is as follows: in Section II, we present the mechanism which generates more realistic networks, i.e. nodes having different fitness which follow a power-law distribution, $\rho(\eta) \propto \eta^{\alpha}$, growth and preferential attachment. Further we present numerical simulations and the method of calculation which are based on monte carlo method and continuum theory respectively employed here. Finally, we determine the exponent γ related to degree distribution, the time evolution of the average number of links, $<k_i>$, the average path length and the clustering coefficient. Section III is devoted to the discussion and results of the topological properties.

2. MODEL AND THEORY

The examples discussed above and others presented in the literature indicate that the incorporation of additional features change the scaling behavior of random growth networks. In particular the BA and fitness model differ markedly in preferential attachment. The first one, $P(i \longleftrightarrow j = N+1) = k_i/\sum_{i=N}^{N} k_i$, obtains $P(k) \propto k^{-3}$. The second one, $P(i \longleftrightarrow j = N+1) = n_i k_i/\sum_{i=N}^{N} n_i k_i$, generates $P(k) \propto k^{-2.25/log(k)}$ if the fitness's variable is uniformly distributed. But in nature we believe that the distribution does not follow this simple distribution. So, we are interested in focusing on the effects of power-law distribution of vertice's quality in the growth dynamics of a network . There are two choices to study these fields, they are (i) $(\alpha > 0)$ it means that a small part of the nodes has short fitness and a major portion of sites has high fitness and (ii) $(-1 < \alpha < 0)$ a little part of nodes has high fitness and a major portion of sites has short fitness. So their distribution is exactly the opposite. In the present paper, we choose the first case $(\alpha > 0)$. It doesn't implicate this case is more realistic than other case. Moreover, both situations converge to the same value of D being also limited by BA model

and Fitness model. We hope with this distribution to better characterize the real aspect of existing correlations in nature and technology.

The complex net was constructed with links which are not direct and weightless. The network grows sequentially by adding a node at each time. It starts with m_0 nodes, each one with fitness η_i that is distributed according to a function given by:

$$\rho = A\eta^{\alpha}, \qquad (1)$$

where A is a constant and the η_i is assigned a value in the range from 0 to 1.

At every time step we add a new node i with fitness η_i. Again this parameter follows the same power-law distribution. We link this node to $m (\leq m_0)$ of the pre-existing nodes in the system. After it, we establish its connections by considering the linking probability which depends on connectivity k_i and the fitness η_i of the node i, and is given by:

$$P(i \longleftrightarrow j = m_0 + 1) = \frac{n_i k_i}{\sum_{i=N}^{N} n_i k_i}. \qquad (2)$$

This procedure is used to include the third site, the fourth site, and so on. The network growth process is sequentially repeated up to the size desired for the system ($N = t + m_0$) where t is the time variable. Note how the rule above privileges the connection between new nodes and those having high nearest neighbors and fitness jointly. The dynamical properties of this model can be addressed using the partial differential equation given by:

$$\frac{\partial k_i(t)}{\partial t} = m \frac{\eta_i k_i}{\sum_j k_j \eta_j}, \qquad (3)$$

where $k_i(t)$ is a continuous real variable and the boundary condition $k_{i=t}(t) = m$. It gives us the temporal dependence of number of links and, more precisely, it tell us how asymptotically it increases with time t/i. The equation 3 describes the rate at which k_i proportionally changes to preferential attachment. The factor m is related to the fact that one new node contributes with m links to the network and the sum in the denominator is over all pre-existed nodes. The solution to equation 3 is given by [4]:

$$k_{\eta_i}(t, t_0) = m \left(\frac{t}{t_0}\right)^{\beta(\eta_i)}, \qquad (4)$$

where t_0 is the time at which the node i was born and the dynamic exponent, $\beta(\eta_i)$, depends on the fitness of the node, i. The dynamic exponent can be written as

$$\beta(\eta_i) = \frac{\eta_i}{D}, \qquad (5)$$

where

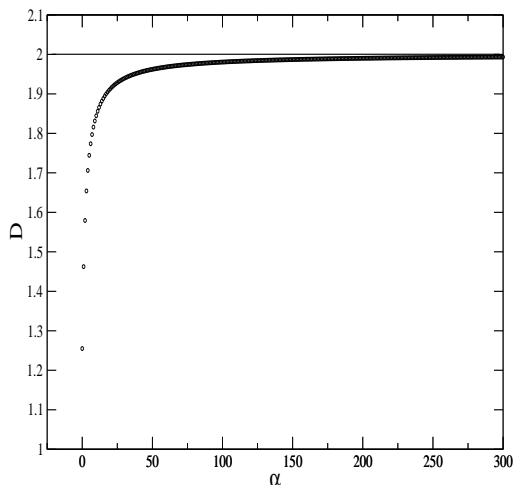$$D = \int d\eta \rho(\eta) \frac{\eta}{1 - \beta(\eta)}. \qquad (6)$$

FIG. 2: Asymptotic behavior of D obtained from equation (8) by varying α.

For a long time, the sum over all existing nodes is

$$< \sum_j \eta_j k_j > \rightarrow Dmt. \tag{7}$$

Note that $\beta(\eta_i)$ depends on D and this depends on $\rho(\eta)$. Then, to calculate the dynamic exponent, it is necessary to solve the integral (6) substituting $\beta(\eta)$ by η/D. After this step, the integral becomes

$$1 = \int_0^{\eta_{max}} d\eta \, \rho(\eta) \frac{1}{\frac{D}{\eta} - 1}, \tag{8}$$

where η_{max} is the highest quality presented in the network. Note that the exponent $\beta(\eta) > 0$ because a node always increases the number of connections with the time evolution. Another condition is that connectivity site, $k_i(t)$, cannot gain more than one link by each time step, t, so $\beta(\eta) < 1$, i.e. the term D/η inside of the integral (7) has a value higher than one. And, the upper limit of constant is $D \leq 2\eta_{max}$, which is obtained from equation (7), and the expression $\sum_j \eta_j k_j \leq \eta_{max} \sum_j k_j = 2mt\eta_{max}$. Finally, we can calculate the value of constant D (associated to the know-how of all the network sites) that only depends on the fitness distribution $\rho(\eta)$ (Eq. ((8)).

Now, we will obtain the exponent γ of degree distribution, $P(k) \sim k^{-\gamma}$ [4, 17]. The probability that a node has a degree $k_i(t)$ smaller than k is:

$$P(k_\eta(t) > k) = P\left[m \left(\frac{t}{t_0} \right)^{\beta(\eta)} > k \right]$$

$$= P\left[t_0 < t \left(\frac{m}{k} \right)^{\frac{D}{\eta}} \right]. \tag{9}$$

Remember that in this model we have a spectrum of dynamics exponents $\beta(\eta)$ and the sites were added at equal time intervals, so the constant probability density is given by

$$P(t_i) = \frac{1}{m_0 + t}. \tag{10}$$

Using the equations 9 and 10, the probability that a site has k links can be written in this form:

$$P(k) = \int_0^{\eta_{max}} d\eta \, \frac{\partial P(k_{\eta(t)} > k)}{\partial k}$$

$$\propto \int_0^{\eta_{max}} d\eta \, \rho(\eta) \frac{D}{\eta} \left(\frac{m}{k} \right)^{\frac{D}{\eta}+1}. \tag{11}$$

Bianconi and Barabási showed the agreement between the prediction (Eqs. 11, 5 e 8) and the simulation for the exponent β and degree distribution $P(k)$ [4].
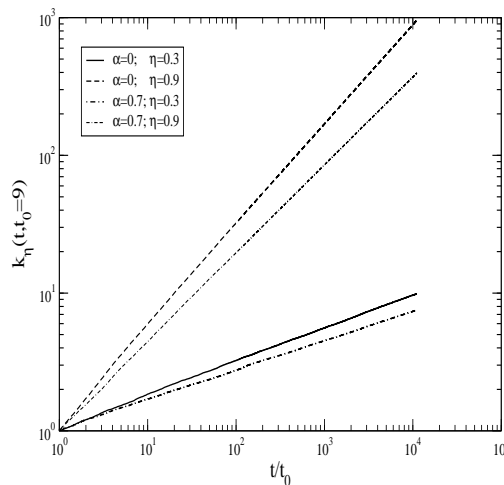

FIG. 3: The symbols are the numerical results of the temporal dependency of the average connectivity for different values of α and η. The plot is on a log-log plot and the results are obtained for networks with 100000 node and 1000 runs.
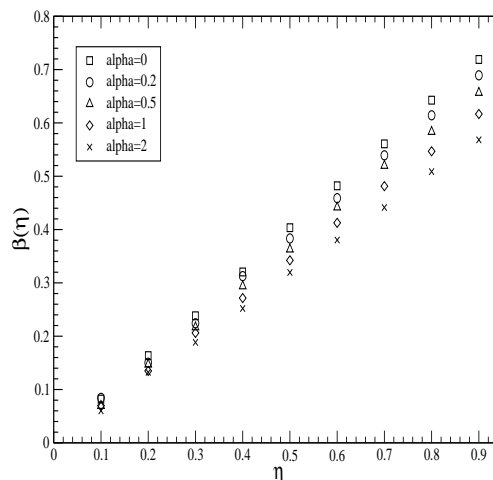

FIG. 4: The symbols are the numerical results of the average connectivity exponent for different values of α related to the measure on node $i = 10$. The results are obtained for networks with 100000 node and 1000 runs.

3. RESULTS AND CONCLUSIONS

In this section we will present our numerical results in order to characterize a scale-free network with power-law distribution of fitness. First of all, to analyze the implications in the quality model using the power-law distribution of fitness, $\rho(\eta) = A\eta^{\alpha}$, where we should substitute ρ in the equation (8). It gives

$$\frac{1}{A} = \int_{D-\eta_{max}}^{D} dy \, \frac{(D-y)^{\alpha+1}}{y}. \tag{12}$$

and its normalization is

$$\int_{0}^{1} \rho(\eta) \, d\eta = 1. \tag{13}$$

Thus, $A = \alpha + 1$ and the fitness distribution diverges for $\alpha < 0$, so we studied only the effects of $\alpha \geq 0$. Solving the integral (12) by numerical methods and varying the parameter α ($\alpha \geq 0$), one may find a crossover between the Barabási *et al.* [3] and Bianconi *et al.* model [4] (Fig. 2). The alternative manner to find the constant D is calculating $< \sum_i \eta_i k_i > /mt \to D$ in the $t \to \infty$ limit. In particular, the cases $\alpha = 0$ and $\alpha \to \infty$ represent the network with uniform distribution of fitness and system without quality factor (or all sites with the same fitness factor $\rho(\eta) = \delta(\eta - 1)$), respectively. It is easy to see the upper limit being the Barabási *et al.* model observing the plot (2). When α goes to infinity, the average fitness goes to unity and $D \to 2$, recovering the BA model ($\beta = 0.5$). Another topologic property (connectivity distribution, average path length and clustering coefficient) must interpolate the BA model and Bianconi *et al.* model (Figs. (1, 5 and 6)).

When we observe the time dependence of the connectivity, $k_{\eta}(t)$ (Figs 3 and 4), an interesting feature of power-law distribution of fitness is found. For a small α factor, the value of fitness is very strong to compete for links, while for a high α factor this difference becomes negligible. The explanation is simple, as the α value increases, the vertices' quality becomes similar, so the term η_i in the preferential attachment becomes only a constant.

In Fig. 1, we show the accumulative connectivity distribution for different values of α and BA model. From this plot, we check that exponent γ varies from 2.255 (Bianconi and Barabási model) to 3 (BA model). The number of hubs and their connectivity decrease as α becomes large. Note that most the networks in nature and technology have exponent γ in similar interval of values studied by us for example: World Wide Web [17], citations [10], telephone [18], and energy landscape [19].

Figure 5 shows the average path length, $< l >$, which depends on α for $m = 2$ and can be compared with the BA model. The average path length, $< l >$, of a network is defined as the number of edges in the shortest path between two nodes averaged over all the pairs of nodes present in the net. We see that, when α increases, $< l >$ behaves like a $< l >$ well-known of BA model. It is explained by the number of hubs in the system. The system with more hubs has a small value of average path length, i.e., the case $\alpha = 0$ has the highest $< l >$ in our study.
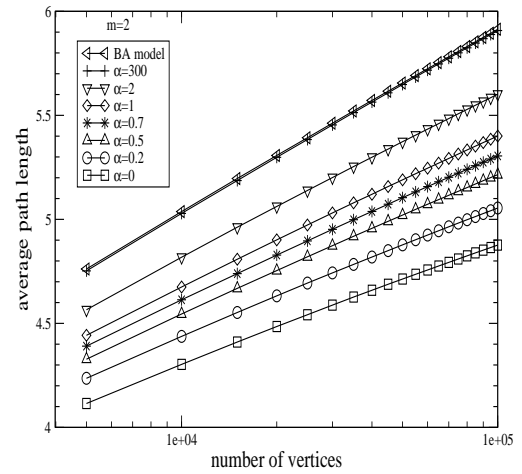

FIG. 5: It shows the average path length in linear-log plot for typical values α. The symbols are the numerical results (80 runs for each size of network).
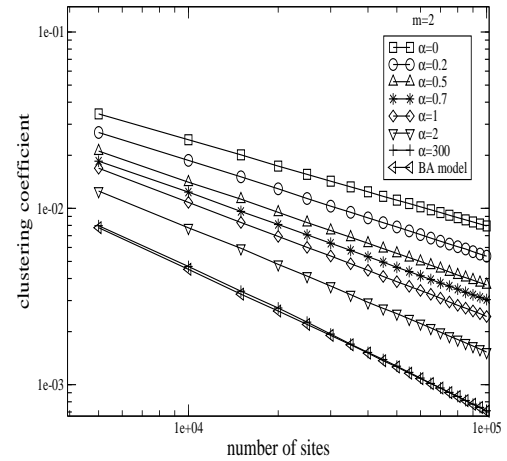

FIG. 6: It shows the clustering coefficient log-log plot for typical values α. The symbols are the numerical results (80 runs for each size of network).

Plot 6 shows the clustering coefficient which depends on α for $m = 2$ and can be compared with the BA model. The clustering coefficient of network characterizes the density of connections close to a node. Note that when α increases, the clustering coefficient becomes similar to Bianconi and Barabási model.

In this paper, we study the effect of the competition between the relevant variable for connectedness when we use a power-law distribution of fitness. In the fitness model, the popular nodes compete with younger nodes when the fitness is an important factor that permits them to obtain more links. By including the exponent α we control the influence of fitness factor of each node in the net. When α is null we recover the Bianconi and Barabási model; when $\alpha \to \infty$, we recover the BA model. The average connectivity $< k_i >$, the average path length $< l >$ and the clustering coefficient are appre-

ciably influenced by the exponent α. When we compare our results with the BA model, we find that the number of links per nodes has risen for all values of α.

[1] P. Erdös and Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).

[2] S.N. Dorogovtsev and J.F.F. Mendes, Physical Review E **62**, 1842 (2000).

[3] A.L. Barabási and R. Albert, Science **286**, 509 (1999).

[4] G. Bianconi and A.-L. Barabási, Europhysics Letters **54**, 436 (2001).

[5] A. Vázquez, cond-mat/0006132 (2000).

[6] D.J.B. Soares, C. Tsallis, A.M. Mariz and L.R. da Silva, Europhysics Letters **70**, 70 (2005).

[7] M.D. de Meneses, D.J.B. Soares, S.D. da Cunha and L.R. da Silva, Progress of Theoretical Physics Supplement **162**, 131 (2006).

[8] R. Albert, H. Jeong and A.-L. Barabási, Nature **401**, 130 (1999).

[9] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley and Y. Aberg, Nature **411**, 907 (2001).

[10] S. Redner, Eur. Phys. J. B **23**, 267 (1998).

[11] P.G. Lind, L.R. da Silva, J.S. Andrade Jr and H.J. Herrmann, Physical Review E **74**, 036117 (2007).

[12] A.A. Moreira, J.S. Andrade Jr and L.A.N. Amaral, Physical Review Letters **89**, 268703 (2002).

[13] P.G Lind, L.R. da Silva, J.S. Andrade Jr and H.J. Herrmann, Europhysics Letters **78**, 68005 (2007).

[14] A.A. Moreira, D.R. Paula, R.N. Costa Filho and J.S. Andrade Jr, Physical Review E **73**, 06501 (2006).

[15] V. Schwmmle, M.C. González, A.A. Moreira, J.S. Andrade Jr and H.J. Herrmann, Physical Review E **75**, 066108 (2007).

[16] S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of Networks* (Oxford University Press, New York, 2003).

[17] R. Albert and A.-L. Barabási, Review of Modern Physics **74**, 47 (2002).

[18] S. Abe and N. Suzuki, Physical Review E **67**, 016106 (2003).

[19] J.P.K. Doye, Physical Review Letters **74**, 238701 (2002).